UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/602,262	06/23/2000	Paul S. Cohen	YOR9-2000-0131-US1	9324

7590 10/22/2003	**EXAMINER**
Anne Vachon Dougherty	LERNER, MARTIN
3173 Cedar Road	
Yorktown Heights, NY 10598	

ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 10/22/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

Status

1) ☒ Responsive to communication(s) filed on <u>22 September 2003</u> .

2a) ☒ This action is **FINAL**.   2b) ☐ This action is non-final.

3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
  closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4) ☒ Claim(s) <u>1 to 10</u> is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5) ☐ Claim(s) _____ is/are allowed.

6) ☒ Claim(s) <u>1 to 10</u> is/are rejected.

7) ☐ Claim(s) _____ is/are objected to.

8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9) ☐ The specification is objected to by the Examiner.

10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11) ☐ The proposed drawing correction filed on _____ is: a) ☐ approved b) ☐ disapproved by the Examiner.

    If approved, corrected drawings are required in reply to this Office action.

12) ☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a) ☐ All b) ☐ Some * c) ☐ None of:

      1. ☐ Certified copies of the priority documents have been received.

      2. ☐ Certified copies of the priority documents have been received in Application No. _____ .

      3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage
        application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

    a) ☐ The translation of the foreign language provisional application has been received.

15) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

DETAILED ACTION

Claim Rejections - 35 USC § 103

1.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

2.   Claims 1, 3 to 5, and 7 to 10 are rejected under 35 U.S.C. 103(a) as being

unpatentable over *Chen* in view of *Braida et al.*

Regarding independent claims 1 and 9, *Chen* discloses a sound-synchronized

video method and system, comprising:

"processing a video signal to generate a video output comprising at least one

time stamped acoustic identification of the content of the audio associated with the

video signal" – codec CD1 separates the digitized video and audio signals into the

digital video and speech components; at the video output of codec CD1, a feature

extraction module FE1 extracts mouth information visemes contain the mouth shape

and mouth location from the decoded video signal; a memory ME1 stores and time

stamps mouth information from the feature extraction module FE1 for phoneme-to-

viseme identification (column 2, lines 5 to 47; column 4, lines 36 to 41: Figure 1);

"processing an audio signal to generate an audio output comprising at least one

[time stamped] acoustic identification of the content of said audio signal" – codec CD1

separates the digitized video and audio signals into the digital video and speech

components; a phoneme recognition module PR1 divides the incoming speech

components into recognizable phonemes; lookup table LT1 maps phonemes into

visemes (column 2, lines 5 to 22; column 4, lines 26 to 35: Figure 1);

"synchronizing the video signal to the audio signal by adjusting at least one of the

signals to align at least one acoustic identification from the video signal with a

corresponding acoustic identification from the audio signal" – video and audio signals

that had become unsynchronized are displayed by synchronizing the video frame to

produce sound synchronized video (column 4, lines 33 to 63: Figure 2).

Concerning independent claims 1 and 9, *Chen* discloses the video signal is time

stamped, but omits time stamping the audio signal. Only one of the audio and video

signals is expressly time stamped in *Chen* because visemes are employed as a

reference to synchronize the signals. However, it is common in the prior art to assign

time stamps to both audio and video data streams for purposes of synchronization to an

absolute time reference. *Braida et al.* teaches a related method and system for

synchronizing video images to speech elements where time stamps are applied to both

audio and video streams. Phone recognition program 44 assigns start and stop times to

digital speech samples 32 (column 6, lines 53 to 58), and digital video images also have

time stamps which are referenced to the same time (column 12, lines 13 to 29). It

would have been obvious to one of ordinary skill in the art to additionally apply time

stamps to the audio signals as taught by *Braida et al.* in the synchronization method and

system of *Chen* for the purpose of providing an absolute time reference for synchronization.

Regarding claim 3, *Chen* discloses phoneme recognition module PR1 produces visemes ("the audio identification") from the audio signal and feature extraction module FE1 extracts corresponding mouth information visemes from lookup table LT1; the output video is applied to display DI1 together with the audio signal and produces lip synchronization (column 2, lines 11 to 38: Figure 1).

Regarding claims 4 and 10, *Chen* discloses a method and system for processing a video image, comprising:

"extracting at least one image from the video signal" – codec CD1 separates the digitized video and audio signals into the digital video and speech components (column 2, lines 6 to 11);

"detecting at least one feature in said at least one image" – a feature extraction module FE1 extracts mouth information visemes contain the mouth shape and mouth location from the decoded video signal (column 2, lines 21 to 39: Figure 1);

"analyzing the parameters of said feature" – mouth deformation module MD1 receives inputs from the video signal and information from the feature extraction module FE1, and visemes from lookup table LT1 (column 2, lines 21 to 39: Figure 1);

"correlating at least one acoustic identification to the parameters of said feature" – a viseme is selected from lookup table LT1 that matches features extracted by feature extraction module FE1 (column 2, lines 21 to 39: Figure 1).

Regarding claims 5 and 7, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes ("articulatory type") rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2).

Regarding claim 8, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes ("articulatory type") rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2); *Chen* discloses feature extraction module FE1 extracts mouth information visemes containing mouth shape ("a facial feature") (column 2, lines 18 to 31).

3.   Claims 2 and 6 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.* as applied to claim 1 above, and further in view of *Basu et al. ('885)*.

Concerning claim 2, *Braida et al.* discloses a Viterbi search for purposes of phone recognition (column 6, lines 59 to 61; column 7, lines 51 to 53), but omits utilizing a Viterbi search for purposes of synchronization. However, it is well known that a Viterbi algorithm is utilized for both recognition and time warping alignment. *Basu et al. ('885)* teaches a method of aligning phonemes and visemes with a Viterbi algorithm. (Column 1, Lines 53 to 67) It would have been obvious to one having ordinary skill in the art to utilize a Viterbi algorithm as suggested by *Basu et al. ('885)* in the

synchronization method and system of *Chen* for the purpose of aligning phonemes and visemes more accurately.

Regarding claim 6, *Chen* discloses speech recognition is at the level of phone groups, corresponding to similar mouth shapes ("articulatory type") rather than individual phonemes (column 3, line 64 to column 4, line 5); similarly, *Braida et al.* processes phones according to context classes (column 8, line 43 to column 9, line 12: Table 2).

Response to Arguments

4.   Applicants' arguments filed 22 September 2003 have been fully considered but they are not persuasive.

Firstly, Applicants argue *Chen* is not synchronizing a "live" video signal to an audio signal, but is overlaying the live signal with stored visemes for a videophone display. This position is traversed.

Neither the claimed invention nor *Chen* says anything about the video signal being "live". The video signal and the audio signal could be recorded and still meet the terms of the claims.

Moreover, *Chen* expressly discloses that the video and audio signals are synchronized. *Chen* notes the audio signal is applied to the output video so as to produce lip synchronization. (Column 2, Lines 37 to 39) Sound synchronized video is produced by modifying mouth area in a current frame. (Column 4, Lines 40 to 44) Thus, the fact that *Chen* synthesizes the video signal from visemes containing generic

mouth shapes in a videophone display does not imply that the reference fails to anticipate the invention as claimed. The feature upon which Applicants rely, that the video image is a "live" image rather than a synthesized image, is not recited in the rejected claims. Although the claims are interpreted in light of the specification, limitations from the specification are not read into the claims. See *In re Van Geuns*, 988 F.2d 1181, 26 USPQ2d 1057 (Fed. Cir. 1993). An additional feature in *Chen*, synthesizing a video signal from stored visemes, still involves a synchronization of a video signal to an audio signal.

Secondly, Applicants argue *Chen* does not teach or suggest the claimed feature of processing a video signal to generate a video output comprising at least one time stamped acoustic identification of the content of the video signal. This position is traversed.

Chen expressly discloses memory ME1 stores and time stamps mouth information from the feature extraction module FE1 for phoneme-to-viseme identification. (Column 2, Lines 40 to 42) The mouth information is stored in a memory of the table LT1 and time stamped for the purpose of phoneme-to-viseme identification. (Column 4, Lines 38 to 41) Thus, *Chen* time stamps the video information of the visemes so as to show how they correspond to the audio sounds of the phonemes. The visemes are video images. Clearly, video information of the visemes is time stamped by an acoustic identification of the phonemes in *Chen*.

Thirdly, Applicants argue *Chen* does not teach or suggest that a video signal be synchronized to the audio signal by adjusting at least one of the signals to align the time

stamped acoustic identification from the video signal with a corresponding acoustic

identification from the audio signal. Instead, Applicants say *Chen* superimposes a

different video signal over the live video signal, the different video signal comprising

visemes that have been fetched from storage. Thus, Applicants conclude *Chen*

expressly teaches a non-synchronous live video signal is covered up in order to appear

synchronous, rather than aligned with the audio signal to actually be synchronous. This

position is traversed.

Chen expressly discloses that the audio and video are synchronized. (Column 2,

Lines 37 to 39; Column 4, Lines 40 to 44) The whole point of the time stamps of *Chen*

is to ensure that the visemes (video output) are synchronized with the phonemes (audio

output). If this were not the case, the video output would not match the audio output.

Thus, *Chen* aligns the video output with the audio output. Although the video images of

Chen are synthetic instead of "live", they are still video signals. The fact that *Chen*

starts with a video image of synthesized visemes for a videophone rather than a live

image of an actual person talking is not dispositive of patentability. A synthesized video

image is still a video image. *Chen* synchronizes the synthesized video image with audio

by time stamping for the purpose of viseme-to-phoneme identification.

Fourthly, Applicants argue *Braida et al.*, which is cited for time stamping, is not

logically combinable with *Chen*, since there would be no reason to time stamp the live

video signal from *Chen* because the latter reference uses stored video/visemes and not

a live video signal. This position is traversed.

The visemes and phonemes of *Chen* still need to be synchronized even though the video image is a synthetic image and not a live image. Applying time stamps to both the video and audio portions of the multimedia program would result in redundancy, where both the video and audio portions are referenced to a common time scale instead of utilizing pointers.

Finally, Applicants argue one would not be motivated to use a Viterbi algorithm for alignment as taught by *Basu et al. ('885)* when alignment of audio to video signals is not done by *Chen*.

It is respectfully maintained this reasoning is circular, and assumes something that is not the case. *Chen* does perform synchronization of video to audio to produce sound synchronized video. Viterbi alignment would be useful for producing a more accurate synchronization of the audio to video by correcting for local variations of the signals.

Therefore, the rejections of claims 1, 3 to 5, and 7 to 10 under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.*, and of claims 2 and 6 under 35 U.S.C. 103(a) as being unpatentable over *Chen* in view of *Braida et al.* as applied to claim 1 above, and further in view of *Basu et al. ('885)*, are proper.

Conclusion

5.   **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later

than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Martin Lerner whose telephone number is (703) 308-

9064. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to

Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Richemond Dorvil can be reached on (703) 305-9645. The fax phone

number for the organization where this application or proceeding is assigned is (703)

872-9306.

Any inquiry of a general nature or relating to the status of this application or

proceeding should be directed to the receptionist whose telephone number is (703) 305-

4700.

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER

ml
10/15/03